Examining Machine Learning for 5G and Beyond through an Adversarial Lens

Examining Machine Learning for 5G and Beyond through an Adversarial Lens

Muhammad Usama[1], Rupendra Nath Mitra[2], Inaam Ilahi[1], Junaid Qadir[1], and Mahesh K. Marina[2]
Information Technology University (ITU), Pakistan[1],
The University of Edinburgh, UK[2]

***Abstract*—Spurred by the recent advances in deep learning to harness rich information hidden in large volumes of data and to tackle problems that are hard to model/solve (e.g., resource allocation problems), there is currently tremendous excitement in the mobile networks domain around the transformative potential of data-driven AI/ML based network automation, control and analytics for 5G and beyond. In this article, we present a cautionary perspective on the use of AI/ML in the 5G context by highlighting the adversarial dimension spanning multiple types of ML (supervised/unsupervised/RL) and support this through three case studies. We also discuss approaches to mitigate this adversarial ML risk, offer guidelines for evaluating the robustness of ML models, and call attention to issues surrounding ML oriented research in 5G more generally.**

***Index Terms*—5G and Beyond Mobile Networks, Adversarial Machine Learning, Security**

I. Introduction

A considerable amount of industry and academic R&D endeavors are currently paving the way toward 5G and Beyond 5G (B5G) networks. 5G networks, unlike their 4G counterparts, are foreseen to be the underpinning infrastructure for a diverse set of future cellular services well beyond mobile broadband to span multiple vertical industries. To flexibly and cost-effectively support diverse use-cases and to enable complex network functions at scale, 5G network design espouses several innovations and technologies such as artificial intelligence (AI) along with software-defined networking (SDN), network function virtualization (NFV), multi-access edge computing (MEC), and cloud-native architecture that are new to the domain of mobile telecommunications.

Technical developments toward 5G and B5G of mobile networks are quickly embracing a variety of deep learning (DL) algorithms as a de facto approach to help tackle the growing complexities of the network problems. However, the well-known vulnerability of the DL models to the adversarial machine learning (ML) attacks can significantly contribute to broadening the overall attack surface for 5G and beyond networks. This observation motivates us to deviate from the on-going trend of developing a newer ML model to address a 5G network problem and, instead, examine the robustness of the existing ML models in relation to the 5G networks under adversarial ML attacks. In particular, we focus on representative use cases for deep neural network (DNN)-driven supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL) techniques in the 5G setting and highlight their brittleness when subject to adversarial ML attacks.

Through this article, we would like to draw the attention of the research community and all stakeholders of 5G and beyond mobile networks to seriously consider the security risks that emerge from the rapid unvetted adoption of DL algorithms across the wide spectrum of network operations, control, and automation, and urge to make robustness of the ML models a criterion before they are integrated into deployed systems. Overall, we make the following two contributions.

1) We highlight that despite the well-known vulnerability of DL models to adversarial ML attacks, there is dearth of critical scrutiny on the impact of the wide-scale adoption of ML techniques on security attack surface of 5G and B5G networks.
2) We bridge the aforementioned gap through a vulnerability study of the DL models in all its major incarnations (SL, UL, and Deep RL) from an adversarial ML perspective in the context of 5G and B5G networks.

II. Background

A. Primer on 5G Architecture

A schematic diagram of the 5G network architecture is depicted in Figure 1. Apart from the user equipment (UE), the 5G system features a cloud-native core network, a flexible and disaggregated radio access network (RAN), and a provision for multi-access edge (MEC) cloud for reduced latency. The RAN comprises of gNodeB (gNB) access nodes, split into distributed and centralized units (DU and CU), to efficiently handle evolved network requirements. The gNB connects to the MEC to significantly reduce the network latency for selected applications by availing edge server computing at the MEC cloud which is close to the radio service cells. For instance, to cater to the ultra-reliable low-latency communication (URLLC) use-case of industry automation, the RAN radio unit along with the DU, CU, and the MEC can be installed onsite. Thus, 5G network architecture enables applications to be deployed remotely (App 3 and App 4) or near the edge (App 1 and App 2), latter when low latency is a requirement. The provision of MEC also reduces the aggregated traffic load on the transport networks responsible for connecting RAN to the core network. The 5G core network (5G-CN) is a cloud-native network that stores subscriber databases and hosts essential virtualized network functions for network operations and management. Although, the network management and control functions are shown to be co-located with the core

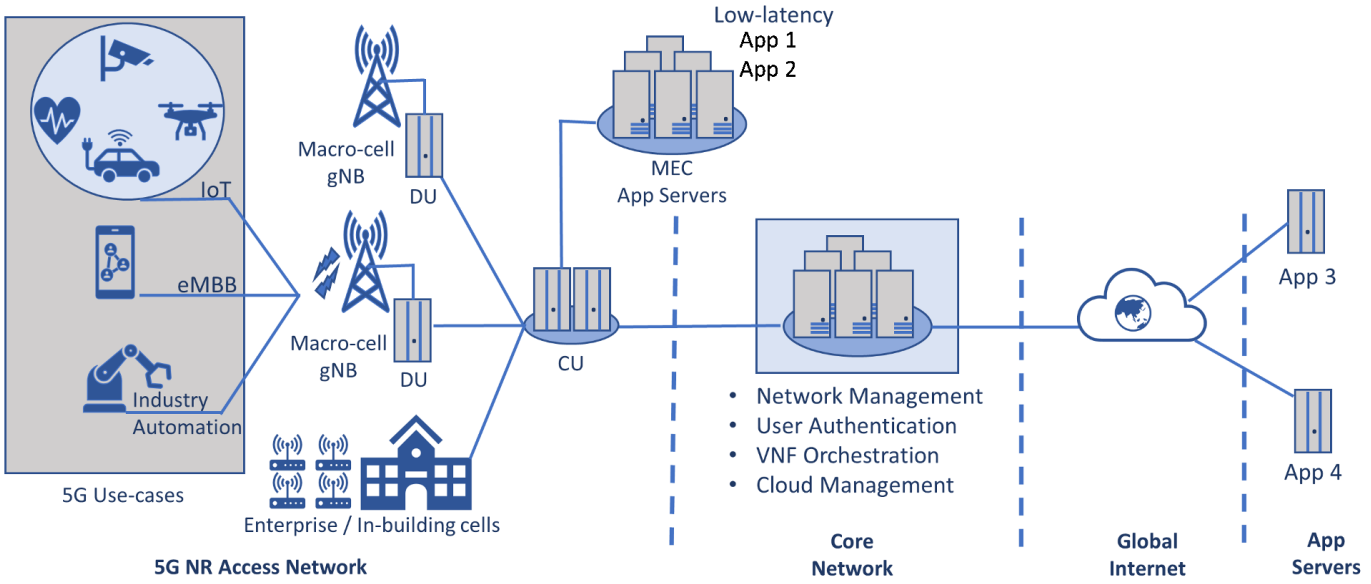

Fig. 1. A schematic diagram of 5G network architecture illustrating the disaggregated RAN architecture with distributed unit (DU) and centralized unit (CU) components; the MEC for improved latency; and the cloud-native core network and system orchestration components.

in the figure, they can be flexibly deployed at the edge as needed.

B. ML in 5G and B5G Networks

A wide spectrum of DL algorithms are being developed for the broad context of wireless communications and 5G networking to deal with problems that are either hard to solve or hard to model. For instance, optimal physical network resource allocation for NFV is an NP-hard problem and so require exponential computational power with increasing system size [1]. Deep RL (DRL)-based solutions are proposed to efficiently address resource allocation problems [2]. Network channel estimation for efficient beamforming is a hard to model problem for which deep neural network (DNN)-based SL solution offers an effective way to tackle it [3]. Moreover, in certain use-cases, conventional expert systems become inappropriate due to real-world constraints, such as limited availability of power, where AI can perform effectively. For instance, deep autoencoder based systems can replace the power-hungry RF chain hardware with small embedded sensor systems enabling them to sustain longer on onboard power supplies. DL algorithms generally outperform the conventional approaches in solving mobile network prediction problems such as physical layer channel prediction by SL, signal detection problems such as recovering transmitted signals from noisy received signals by UL, and optimization problems like resource allocation by RL.

III. Widened Attack Surface in ML-Driven 5G and B5G Networks

The security of the 5G networks is receiving great deal of attention (e.g., [4]), but there is very limited focus on the security of 5G and B5G networks in the face of adversarial ML threat [5]. In this section, we briefly introduce the adversarial ML in general, and subsequently outline the adversarial ML risks in 5G and B5G networks.

A. Overview of Security Attacks on ML

The vulnerability of the ML algorithms, especially the DL models, to the adversarial attacks is now well-established, where adversarial inputs are small carefully-crafted perturbations in the test data built for fooling the underlying ML model into making wrong decisions. An adversary can often successfully target an ML model with no knowledge of the model (*black-box attack*), or some knowledge (*grey-box attack*), or full knowledge (*white-box attack*) of the target model. An adversary can attack the model during its training phase and in its testing phase as well. The training phase attacks are known as *"poisoning attacks"* and the test time attacks are known as *"evasion attacks"*. Evasion attacks are commonly known as adversarial attacks in the literature [6].

More formally, an *adversarial example* x^* is crafted by adding a small indistinguishable perturbation δ to the test example x of a trained ML classifier $f(.)$ where δ is approximated by the nonlinear optimization problem provided in equation 1, where t is the class label.

$$x^* = x + \arg\min_{\delta_x}\{\|\delta\| : f(x+\delta) = t\} \quad (1)$$

In 2013, Szegedy et al. [7] observed the discontinuity in the DNN's input-output mapping and reported that DNN is not resilient to the small changes in the input. Following on this discontinuity Goodfellow et al. [8] propose a gradient-based optimization method for crafting adversarial examples. This technique is known as *fast gradient sign method* (FGSM). Papernot et al. [9] craft adversarial perturbation using a saliency map-based approach on the forward derivatives of DNN. This

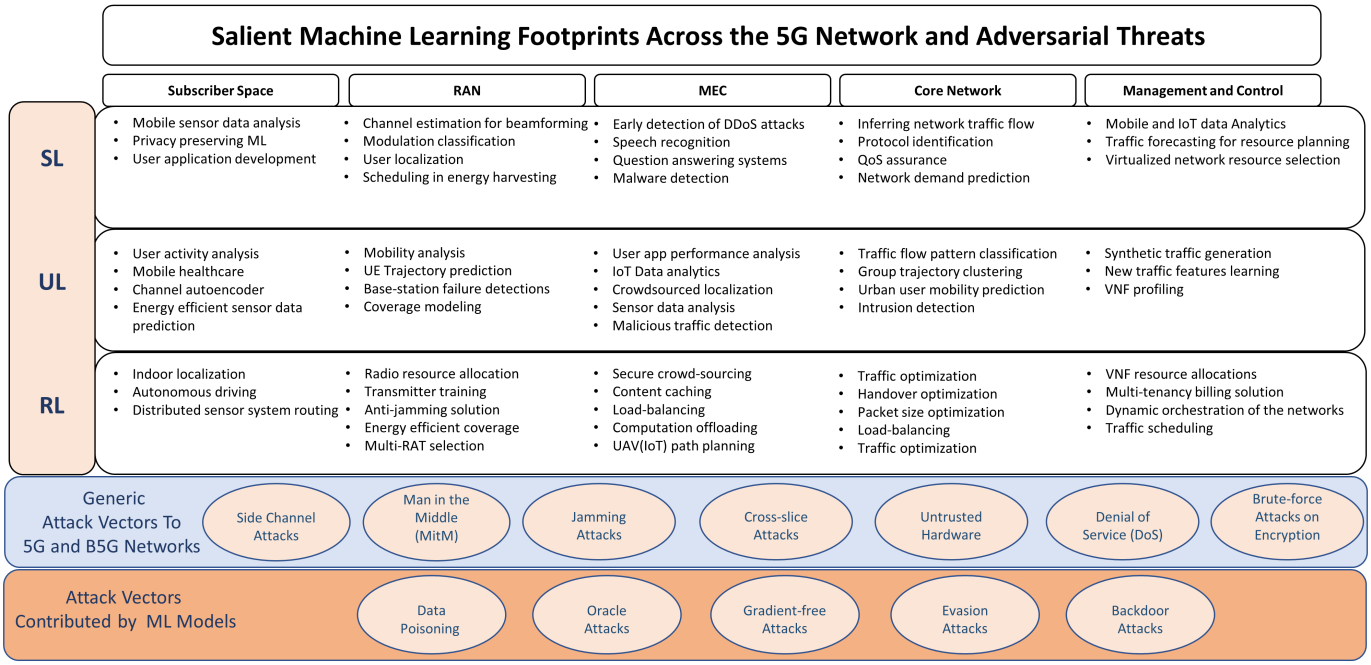

Fig. 2. Applicability of ML across the 5G network architecture and a depiction of how ML models contribute to significantly enhance the attack vectors beyond the traditional security risks [4] with new adversarial ML risks.

approach is known as *Jacobian saliency map based attack* (JSMA). Carlini et al. [10] crafted three different adversarial attacks using three different distance matrices (L_1, L_2, and L_∞). More details about adversarial ML attacks are described in [6].

It is important, however, to note that the adversary does not need to have access to training or test datasets. Instead, adversarial examples can also be generated using query efficient gradient-based techniques [11], zeroth order optimization techniques [6], and generative models [12]. In such methods, the adversary uses query-response pairs to craft such adversarial examples (inputs) and mislead the ML model. Such pairs are not necessarily part of either training or testing datasets, therefore, adversarial examples are not just the result of an input data security issue.

B. Added Threat from Adversarial ML for 5G and Beyond

Figure 2 illustrates network problems from different network segments of 5G, namely user devices, RAN, MEC, core networks, and the network management and control layer that have recently attracted ML-based solutions from all the three categories of ML. However, in light of the above discussion in §III-A, the DL-powered ML models gaining popularity for 5G and B5G networks are vulnerable to the adversarial attacks thereby further aggravating the security risks of future generations of mobile networks.

To show the feasibility of adversarial ML attacks on 5G systems we take three well-known ML models—one from each of the three ML families of algorithms (UL, SL, and DRL)—from wireless physical layer operations relevant to 5G and B5G context and show the vulnerability that naive use of ML brings to future mobile networks. We choose all the three ML models for our case studies from the physical layer network operations because of the maturity of ML-research in the context of AI-driven 5G networking and the availability of open-sourced ML models backed up with accessible datasets[1].

IV. Three Case Studies Highlighting Adversarial ML Risk for 5G and Beyond

A. Attacking ***Supervised ML-based*** 5G Applications

Automatic modulation classification is a critical task for intelligent radio receivers where the signal amplitude, carrier frequency, phase offsets, and distribution of noise power are unknown variables to the receivers subjected to real-world frequency-selective time-varying channels perturbed by multipath fading and shadowing. The conventional maximum-likelihood and feature-based solutions are often infeasible due to the high computational overhead and domain expertise that is required. To make modulation classifiers more common in modern 5G and B5G networked devices, current approaches deploy DL to build an end-to-end modulation classification systems capable of automatic extraction of signal features in the wild.

We pick a convolutional neural network (CNN)-driven SL-based modulation classification model in this case study to illustrate the added dimension of vulnerability introduced in the networks by it. We use the well-known GNU radio ML RML2016.10a dataset that consists of 220000 input examples of 11 digital and analog modulation schemes (AM-DSB, AM-SSB, WBFM, PAM4, BPSK, QPSK, 8PSK, QAM16, QAM64, CPFSK, and GFSK) on the signal to noise ratio (SNR) ranging

[1]https://mlc.committees.comsoc.org/research-library/

from -20 dB to 18dB [13]. However, we exclude the analog modulation schemes from our study and consider only the eight digital modulations from the data set because from 2G onward all mobile wireless standards are strictly digital communications. Figure 3 depicts the classification performance of the CNN model in the multi-class modulation classification for the signals between -20dB to 18dB of SNR.

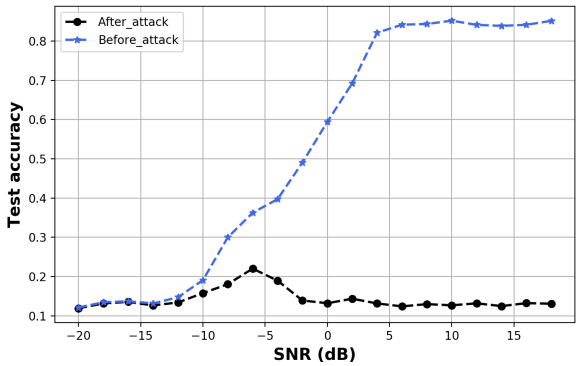

Fig. 3. Accuracy of the CNN-based automatic modulation classifier before and after the adversarial ML attack. *A clear drop in the accuracy of the classifier with the increasing SNR indicates the success of the adversary in compromising the integrity of the modulation classifier that is seen as viable in the 5G and B5G networks.*

To show the feasibility of an adversarial ML attack on the CNN-based modulation classifier we make the following assumptions:

- We consider the *white-box* attack model where we assume that the adversary has a complete knowledge about the deployed modulation classifier.
- *Goal of the adversary* is to compromise the integrity of the CNN classifier leading to a significant decay in the classification accuracy which is the measure of the *success of the adversary*.

To craft the adversarial examples to fool the CNN classifier, we use the Carlini & Wagner (C&W) attack [10] for each modulation class by minimizing the L_2 norm on the perturbation δ, such that when the perturbation δ is added to the input x and sent to the CNN-based modulation classifier C it misclassifies the input x. More details on the C&W attack are available in [10]. The performance of the CNN-based modulation classifier before and during the adversarial attack is depicted in Figure 3. A distinct drop in the accuracy of the modulation classification after the adversarial attacks indicates the brittleness of deep supervised ML in 5G and B5G applications. Moreover, our results show that the adoption of unsafe DL models in the physical layer operations of the 5G and B5G networks can make the air-interface of the future networks vulnerable to adversarial ML attacks.

*B. Attacking **Unsupervised ML-based** 5G Applications*

In 2016, O'Shea et al. proposed the idea of channel autoencoders which is an abstraction of how an end-to-end radio communication module functions in real-world wireless systems [15]. Such a deep autoencoder-based communication model is seen as a viable alternative to the dedicated radio hardware in the future 5G and beyond networks [16]. Figure 4(a) depicts the conceptual design of the channel autoencoder that we choose as a deep UL model for this case study. We assume the model is subjected to an additive white Gaussian noise (AWGN) channel and apply the parameter-configurations provided in [14]. To perform the adversarial ML attack on the channel autoencoder we consider the following threat model and compare the performance of the model with and without attack.

- We assume a *white-box* setting, where the adversary has complete knowledge of the deployed ML model. We further assume that the autoencoder learns a broadcast channel. The proposed adversarial attack on channel autoencoder can be converted into a black-box adversarial attack, where the adversary has zero knowledge of the target ML model, by following the surrogate model approach provided in [11].
- The *goal of the adversary* is to compromise the integrity of channel autoencoder and the *success of the adversary* is measured by the elevated block error rate (BLER) with improving SNR per bit (E_b/N_0).

We take the following two-step data-independent approach to craft adversarial examples for the channel autoencoder:

1) Sample the Gaussian distribution randomly (because the channel is AWGN) and use it as an initial adversarial perturbation δ;
2) Maximize the mean activations of the decoder model when the input of the decoder is the perturbation δ.

This produces maximal spurious activations at each decoder layer and results in the loss of the integrity of the channel autoencoder. Figure 4(b) shows the performance of the model before and under the adversarial attack. Moreover, the figure suggests that adversarial ML attack often outperforms the traditional jamming attacks.

Since the idea of channel autoencoder in a wireless device is to model the on-board communication system as an end-to-end optimizable operation, the adversarial ML attacks on channel autoencoder show that the application of unsupervised ML in the 5G mobile networks increases its vulnerability to adversarial examples. Hence, we argue that deep UL-based 5G networked systems and applications need to be revisited for their robustness before being integrated into the 5G, IoT, and related systems.

*C. Attacking **Reinforcement ML-based** 5G Applications*

In the final case study, we performed the adversarial ML attacks on an end-to-end DRL autoencoder with a noisy channel feedback system [17]. Goutay et al. [17] take the same architecture we consider in the previous case study §IV-B and add a noisy feedback mechanism to it, as shown in Figure 5(a). The end-to-end training procedure involves:

1) The RL-based transmitter training by a policy gradient theorem [17] to ensure that the intelligent transmitter learns from the noisy feedback after a round of communication.

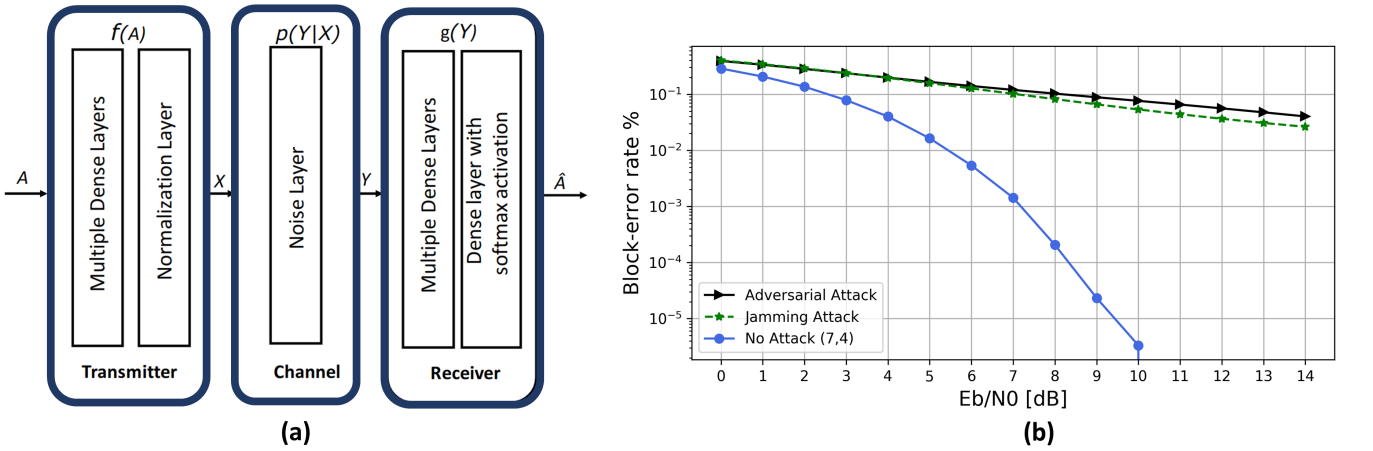

Fig. 4. (a) Architecture of channel autoencoder for 5G and future networks proposed in [14]; (b) Performance of the channel autoencoder before and under the adversarial ML attack and traditional jamming attack. *The Block Error Rate (BLER) versus E_b/N_0 curves indicates that adversarial ML attack does not only deteriorate the model's performance but also leads to similar or worse performance than with a known jamming attack.*

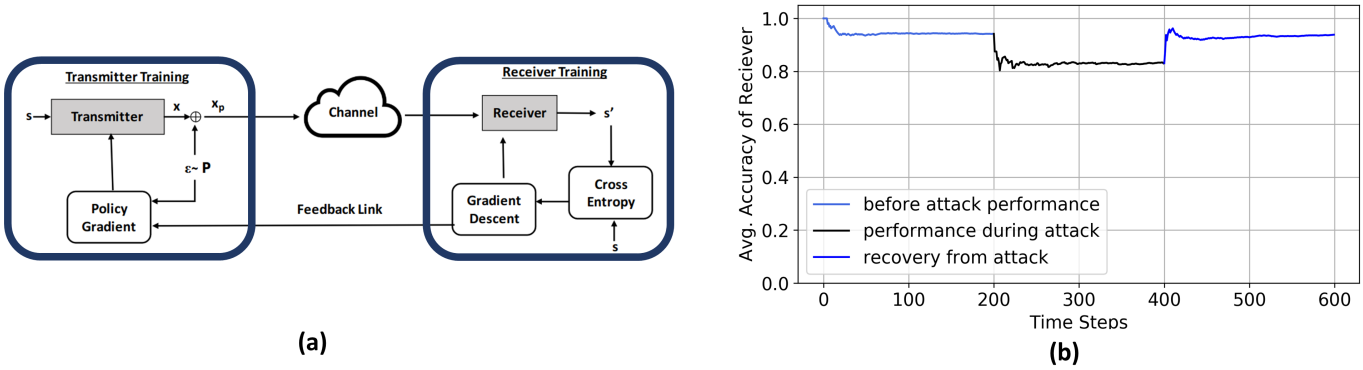

Fig. 5. (a) Architecture of DRL-based channel autoencoder with noisy feedback for 5G and B5G networks proposed in [17]; (b) Performance of DRL autoencoder with noisy feedback before, during, and after the adversarial ML attack. *A clear drop in the performance of the receiver during the attack indicates the success of the adversary in compromising the DRL autoencoder-based end-to-end communication system in future mobile networks.*

2) SL model-based receiver training to train the receiver as a classifier.

More details on the design and training procedure are available in [17]. The considered threat model for this case study is given as:

- We choose a realistic *black-box* settings where the adversary does not know the target model. We also assume that the adversary can perform an adversarial ML attack for "n"-time steps.
- The *goal of the adversary* is to compromise the performance of the DRL autoencoder with noisy feedback for a specific time interval. The *success of the adversary* is measured by the degradation in the decoder's performance during the attack interval.

We exploit the transferability property of the adversarial examples, which states that adversarial examples compromising an ML model will compromise other ML models with high probability if the underlying data distribution is same between two victim models. So we transfer the adversarial examples crafted in case study (§IV-B) and measure the average accuracy of the receiver. We run the DRL autoencoder with a noisy feedback system for 600-time steps (one time-step is equal to one communication round) and perform the adversarial attack between 200 to 400-time step window. We transfer 200 successful perturbations from the previous case study (§IV-B). Figure 5(b) shows the performance of the receiver (decoder) of the DRL autoencoder. It is evident that the performance of the receiver degrades from 95% to nearly 80% during the adversarial attack window.

Our results, as presented in this section, confirm the feasibility of adversarial ML attacks on DL-based applications from all the three types of ML algorithms that are prevalent in the 5G network systems, and highlight the additional threat landscape emerges due to the integration of vulnerable DL models to the 5G and B5G networks.

V. Discussion

A. Towards Robust ML-Driven 5G and Beyond Networks

Robustness against adversarial ML attacks is a very challenging problem. We first note that there does not exist much work on the recommendations and guidelines for evaluating the robustness of ML in 5G applications. Moreover, to date,

there does not exist a defense that ensures complete protection against adversarial ML attacks. In our previous works [6], [18], we have performed an extensive survey of the adversarial ML literature on robustness against adversarial examples, and showed that nearly all defense mechanisms proposed in the literature take one of the following three approaches:

1) modifying data (e.g., adversarial training, feature squeezing, input masking);
2) auxiliary model addition (e.g., generative model addition, ensemble defenses);
3) modifying model (e.g., defensive distillation, model masking, gradient regularization).

Although, our results in 5G related use-cases presented in section IV indicate that the representative ML-based 5G applications from physical layers are vulnerable to the adversarial ML attacks, the threat models exploit the underlying vulnerability inherent to known DL models in general. For instance, we were able to attack the DRL auto-encoder by exploiting the fact of transferability which is the root-cause that enables a same perturbation to fool multiple models. Thus, we draw attention to the security landscape of 5G and B5G widening further from adoption of a plethora of DL-driven components, substantiated through results from three specific use cases related to 5G physical layer.

B. Recommendations for designing and evaluating defenses against adversarial ML attacks

1) Designing a defense: Designing a defense against adversarial examples is a very challenging task. Many approaches for defending against these attacks are available in the literature but these techniques are shown ineffective against newer variations of the attacks [6]. The following are a few recommendations for designing a defensive intervention against adversarial examples.

- A generic defense that can defend against any type of adversarial attack is not possible. So the first logical step is to understand the threat model of the system for which the defensive intervention is needed.
- In many cases, the adversarial examples are generated/sampled from a distribution similar to the legitimate data. A preemptive data generation process (by using generative models) and aggressive labeling (labeling the preemptively generated examples as false positives) can improve the odds of detecting many adversarial attacks. In our previous work [12], we have shown that this procedure can help in making a better defense.
- Deploy all known procedures from the literature that is in line with the threat model.
- Always design defenses considering adaptive adversaries.

2) Evaluating a defense: In the following, we have provided a few important evaluation guidelines for evaluating the ML-based 5G applications against adversarial ML attacks. These insights are extracted from the Carlini et al. [19] and our previous works [6], [20].

- Many defenses are available in the literature against adversarial attacks but these defenses are limited by the design of the application. Using them without considering the threat model of ML-based 5G applications can create a false sense of security. So, for ML-based 5G applications, threat models must clearly state the assumptions taken, type of the adversary, and the metrics used for evaluating the defense.
- Always test the defense against the strongest known attack and use it as a baseline. Evaluating for an adaptive adversary is also necessary.
- Evaluate the defense procedure for gradient-based, gradient-free, and random noise-based attacks[2].
- Clearly state the evaluation parameters (accuracy, recall, precision, F1 score, ROC, etc.) used in evaluating/validating the defense, and always look for a change in the false positive and false negative scores.
- Evaluation of the defense mechanism against out-of-distribution examples and transferability-based adversarial attacks is very important.

Although these recommendations and many others in [6], [18]–[20] can help in designing a suitable defense against adversarial examples but this is still an open research problem in adversarial ML and ripe for investigation for ML-based 5G applications.

C. Beyond Vulnerability to Adversarial ML Attacks

Apart from the vulnerability of the ML models to the adversarial ML attacks, we underline the following drawbacks that call into question the possibility of ML-driven solutions getting integrated into the real-world 5G networks any time soon.

1) Lack of real-world datasets: Due to the dearth of openly available real network data from the telecom operators, a large amount of ML research in the telecom domain still largely depends on simulated/experimental data that often falls short of truly representing real-world randomness and variations. Thus, current state-of-the-art ML models in telecommunication applications are not yet ready to replace the domain-knowledge based expert systems currently in operation.

2) Lack of explainability: In ML studies, the accuracy of a model comes at the cost of explainability. The DL models are highly accurate in providing output but lack an explanation of why a particular output is achieved. Explanation of a decision taken often would be a critical requirement in the 5G and B5G network settings, especially because many critical services such as transport signaling, connected vehicles, and URLLC are expected to be realized over the 5G infrastructure.

3) Lack of operational success of ML in real-world mobile networks: A plethora of ML models exist in the mobile networking literature but use of ML models in operational mobile networks currently is still quite limited. When we perform attacks on the ML models running under the ideal environment, simulated or in favorable lab conditions, and still, the victim models cannot withstand the adversarial attacks, as demonstrated through our case studies. In real-world mobile networks, the ML models need to be deployed and stay functional under unforeseen random environments, leaving

[2]https://www.robust-ml.org/

them more vulnerable to adversarial attacks that are beyond what they are designed to be robust against.

VI. Conclusions

Security and privacy are uncompromising necessities for modern and future global networks standards such as 5G and Beyond 5G (B5G), and accordingly fortifying it to thwart attacks and withstand the rapidly evolving landscape of future security threats is of vital importance. This article specifically highlights that the unvetted adoption of deep learning driven solutions in 5G and B5G networking gives rise to security concerns that remain unattended by the 5G standardization bodies, such as the 3GPP. We argue this is the right time for cross-disciplinary research endeavors considering ML and cybersecurity to gain momentum, and enable secure and trusted future 5G and B5G mobile networks for all future stakeholders. We hope that our work will motivate further research towards "telecom-grade ML" that is safe and trustworthy enough to be incorporated into 5G and beyond 5G networks, thereby power intelligent and robust mobile networks supporting diverse services including mission-critical systems.

References

[1] Aun Haider, Richard Potter, and Akihiro Nakao. Challenges in resource allocation in network virtualization. In *20th ITC specialist seminar*, volume 18. ITC, 2009.

[2] Xenofon Foukas, Mahesh K Marina, and Kimon Kontovasilis. Iris: Deep reinforcement learning driven shared spectrum access architecture for indoor neutral-host small cells. *IEEE Journal on Selected Areas in Communications*, 37(8):1820–1837, 2019.

[3] Hongji Huang, Jie Yang, Hao Huang, Yiwei Song, and Guan Gui. Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system. *IEEE Transactions on Vehicular Technology*, 67(9):8549–8560, 2018.

[4] Ijaz Ahmad, Shahriar Shahabuddin, Tanesh Kumar, Jude Okwuibe, Andrei Gurtov, and Mika Ylianttila. Security for 5G and beyond. *IEEE Communications Surveys & Tutorials*, 21(4):3682–3722, 2019.

[5] Yalin E Sagduyu, Yi Shi, Tugba Erpek, William Headley, Bryse Flowers, George Stantchev, and Zhuo Lu. When wireless security meets machine learning: Motivation, challenges, and research directions. *arXiv preprint arXiv:2001.08883*, 2020.

[6] Adnan Qayyum, Muhammad Usama, Junaid Qadir, and Ala Al-Fuqaha. Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward. *IEEE Communications Surveys & Tutorials*, 22(2):998–1026, 2020.

[7] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[8] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples (2014). *arXiv preprint arXiv:1412.6572*.

[9] Nicolas Papernot, Patrick McDaniel, Somesh Jha, Matt Fredrikson, Z Berkay Celik, and Ananthram Swami. The limitations of deep learning in adversarial settings. In *Security and Privacy (EuroS&P), 2016 IEEE European Symposium on*, pages 372–387. IEEE, 2016.

[10] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *Security and Privacy (SP), 2017 IEEE Symposium on*, pages 39–57. IEEE, May 2017.

[11] Muhammad Usama, Adnan Qayyum, Junaid Qadir, and Ala Al-Fuqaha. Black-box adversarial machine learning attack on network traffic classification. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 84–89. IEEE, 2019.

[12] Muhammad Usama, Muhammad Asim, Siddique Latif, Junaid Qadir, and Ala Al-Fuqaha. Generative adversarial networks for launching and thwarting adversarial attacks on network intrusion detection systems. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 78–83. IEEE, 2019.

[13] Timothy J O'Shea and Nathan West. Radio machine learning dataset generation with GNU radio. In *Proceedings of the GNU Radio Conference*, volume 1, 2016.

[14] Timothy O'Shea and Jakob Hoydis. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

[15] Timothy J O'Shea, Kiran Karra, and T Charles Clancy. Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention. In *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 223–228. IEEE, 2016.

[16] Hilburn, Ben and O'Shea, Timothy J and Roy, Tamoghna and West, Nathan. DeepSig: Deep Learning for Wireless Communications, 2018. Retrieved July, 2020 from https://developer.nvidia.com/blog/deepsig-deep-learning-wireless-communications/.

[17] Mathieu Goutay, Fayçal Ait Aoudia, and Jakob Hoydis. Deep reinforcement learning autoencoder with noisy feedback. In *2019 17th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, (WIOPT)*, 2019.

[18] Inaam Ilahi, Muhammad Usama, Junaid Qadir, Muhammad Umar Janjua, Ala Al-Fuqaha, Dinh Thai Hoang, and Dusit Niyato. Challenges and countermeasures for adversarial attacks on deep reinforcement learning. *arXiv preprint arXiv:2001.09684*, 2020.

[19] Nicholas Carlini, Anish Athalye, Nicolas Papernot, Wieland Brendel, Jonas Rauber, Dimitris Tsipras, Ian Goodfellow, Aleksander Madry, and Alexey Kurakin. On evaluating adversarial robustness. *arXiv preprint arXiv:1902.06705*, 2019.

[20] Muhammad Usama, Junaid Qadir, Ala Al-Fuqaha, and Mounir Hamdi. The adversarial machine learning conundrum: Can the insecurity of ml become the achilles' heel of cognitive networks? *IEEE Network*, 34(1):196–203, 2019.